

Commentary

# Hospitalists Are Already Using AI—Why Implementation Will Determine Its Impact

Anna Maw\*, MS, MD; Aakriti Pandita\*, MD; Marisha Burden, MBA, MD

Division of Hospital Medicine, University of Colorado School of Medicine, Aurora, CO, United States

\*these authors contributed equally

**Corresponding Author:**

Aakriti Pandita, MD  
Division of Hospital Medicine  
University of Colorado School of Medicine  
Leprino Building, 4th Floor, Mailstop F-782, 12401 E 17th Avenue  
Aurora, CO 80045  
United States  
Phone: 1 720 848 4289  
Email: [aakriti.pandita@cuanschutz.edu](mailto:aakriti.pandita@cuanschutz.edu)

**Related Article:**

Comment on: <https://www.jmir.org/2026/1/e85973>

## Abstract

The adoption of artificial intelligence (AI) into clinical practice is accelerating, outpacing the development of organizational guidance, training, and governance. A recent study indicated that two-thirds of hospitalists are using AI, particularly large language model (LLM)–based platforms, in their clinical work. However, as with prior disruptive health technologies, adoption alone does not ensure meaningful improvement in care. Drawing on lessons from electronic health record implementation, we argue that AI's ultimate impact will be determined not by use rates, but by implementation quality and fit. Poorly implemented digital tools have been shown to increase clinician workload and burnout, despite their intended benefits. Early evidence on LLM-based diagnostic AI further underscores this risk: clinical-decision making supported by AI may be suboptimal when integration, training, and workflow design are inadequate. To provide value, AI tools must be thoughtfully embedded into clinical reasoning processes through evidence-informed training, intentional workflow design, and supportive organizational culture. As AI technologies are rapidly adopted, three priorities come into focus: training clinicians on AI inputs and interpreting outputs, applying implementation science frameworks for AI deployment in clinical environments, and establishing strategies for ongoing evaluation of the impact of AI tools over time. Implementation science frameworks offer practical guidance to assess workflow integration, training needs, infrastructure, and potential unintended consequences that can then inform adaptation of implementation strategies to enhance contextual fit. In parallel, learning health system infrastructure can enable continuous monitoring and iterative adaptation using routinely collected clinical and workflow data that reflect the value of the intervention across the quintuple aim of clinical outcomes, health equity, cost, and patient and clinician experience. AI adoption in hospital medicine is likely inevitable. Its ability to advance the quintuple aim will depend on how effectively these tools are implemented, supported, evaluated, and adapted in practice.

*J Med Internet Res* 2026;28:e97419; doi: [10.2196/97419](https://doi.org/10.2196/97419)

**Keywords:** artificial intelligence; generative AI; hospital medicine; clinical decision support systems; health care technology adoption; implementation science

The integration of artificial intelligence (AI) into clinical practice is accelerating, with clinicians adopting AI tools faster than health systems can guide and govern their use, often without structured training or guidance to interpret outputs or manage potential risk. In a recent study by Bagla and colleagues [1], two-thirds of hospitalists reported

using AI platforms in clinical practice, most commonly large language model (LLM)–based tools, and were doing so largely in the absence of health system integration. Hospitalists reported using these tools primarily to support clinical decision-making, including generating differential diagnoses and management options, with a strong preference for

medical-specific platforms over general-purpose applications. Yet as with any disruptive health technology, adoption alone should not be considered success. Prior experience has shown that widespread uptake of technology does not guarantee progress toward the quintuple aim of optimizing patient outcomes, health equity, cost, and patient and clinician experience [2]. The work by Bagla et al [1] provides an important foundation by highlighting how AI is already being used in practice, with broader applications and tools likely to continue expanding. The critical next question is whether and under what conditions these patterns of use translate into meaningful improvements across the quintuple aim, particularly as health systems begin to consider how these tools should be implemented, supported, and governed.

Whether clinician-initiated (as with many current AI tools) or system-led (as with electronic health record [EHR] implementation), health technologies often fall short when implementation, training, and workflow redesign lag behind adoption. The history of EHR implementation provides a cautionary example of adopting technology without careful attention to implementation and contextual fit. While EHRs have contributed to important gains in standardization and safety, they have also introduced risks, including interoperability challenges, clinician inefficiencies and cognitive burden, and even downstream patient safety concerns [3]. Importantly, these harms largely reflect design and implementation failures rather than what the technology itself can offer. As emerging AI technologies are rapidly adopted, three priorities come into focus: training clinicians on prompt engineering and interpreting outputs, applying implementation science frameworks for AI deployment in clinical environments, and establishing strategies for ongoing evaluation of the impact of AI tools over time.

As health systems move toward integration of various AI technologies (including LLM-based generative AI solutions and beyond), emerging evidence from randomized controlled trials highlights how implementation and training critically shape their impact. For example, when comparing diagnostic performance of LLMs across three groups—clinicians alone, clinicians with AI assistance, and AI alone—Goh and colleagues [4] reported that AI assistance did not improve diagnostic reasoning of clinicians and AI alone outperformed clinicians. This does not necessarily imply that clinicians should be replaced; rather, it reflects a critical gap in implementation. Simply adding AI to clinical practice is insufficient. Without evidence-informed training and structured approaches to use, clinicians may fail to appropriately prompt, interpret, contextualize, and act on AI outputs, limiting the clinical benefits that could be achieved. Additional evidence from randomized controlled trials further demonstrates that it is not only *whether* AI is used, but *how* and *when* it is integrated into clinical reasoning that shapes outcomes. Everett et al [5] demonstrated that when clinicians received targeted training and used intentionally designed AI-first or AI-second workflows, diagnostic accuracy improved markedly, highlighting that *how* AI is integrated influences outcomes. Similarly, early studies of

AI-enabled documentation tools, such as AI scribes, show promise but also demonstrate variable benefit across care settings and workflows [6].

To advance the quintuple aim, AI use in clinical care must be approached as more than simply deploying a new tool, with careful attention to how it is introduced, supported, and used in practice [7]. This is particularly important for LLM-based platforms being used to support clinical decision-making, whose outputs vary by prompt input and can include overly confident, sometimes erroneous output that could negatively impact clinician decision-making. Implementation science frameworks provide practical scaffolding for operational decision-making, particularly as AI tools are rapidly deployed into clinical environments. For instance, the pragmatic robust implementation and sustainability model (PRISM) can help health system leaders systematically assess key factors, such as workflow integration, organizational culture, infrastructure (such as training), and external pressures, and anticipate and measure how these elements influence outcomes and unintended consequences [8]. In practice, this means moving beyond “Does it work?” to more actionable questions: “Where in our workflows should this tool be used?” “Who needs training?” “What could go wrong?” “How will we monitor and improve performance over time?” and “What are the impacts on patients, the workforce, and the health system?” At the same time, responsible use extends beyond organizational strategy to individual clinicians who are already incorporating AI tools into their practice. Clinicians must understand the strengths and limitations of the AI tools they use clinically, and how the outputs of these tools influence their decision-making. Clinicians must also ensure that their tool use aligns with Health Insurance Portability and Accountability Act (HIPAA) requirements and organizational policies.

Finally, to understand whether training and deployment strategies are working in practice, implementation of AI tools at scale will require ongoing health system evaluation and adaptation rather than one-time deployment. Learning health system infrastructure, which allows system leaders to leverage data collected through routine clinical care, can enable continuous, near-real-time monitoring of process and outcome measures, making timely iterative improvement feasible at scale [9]. Fortunately, much of these data already exist within the EHR, from audit logs that reveal workflow patterns to clinical data that track patient outcomes [10]. However, measurement must extend beyond what is easily captured: periodic assessments of clinician experience and workflow burden are critical, as are economic evaluations to ensure that AI tools deliver on their investment.

Ultimately, AI adoption in hospital medicine (and beyond) will only continue, but its impact on critical outcomes remains uncertain. The central challenge is no longer whether these tools are adopted but whether they are implemented in ways that meaningfully improve outcomes for patients, clinicians, and health systems.

---

## Acknowledgments

The authors used Microsoft CoPilot and ChatGPT versions 5 and 5.3 for editing of original content to improve readability. All information and materials in the manuscript are original.

---

## Funding

The authors declared no financial support was received for this work.

---

## Authors' Contributions

All authors contributed to conceptualization, writing, review, and editing of the manuscript. All authors reviewed and approved the final manuscript.

---

## Conflicts of Interest

MB reports funding from the Agency for Healthcare Research and Quality, the National Institute for Occupational Health and Safety, a University of Colorado Innovations digiSPARK award, and the American Medical Association not related to this work. MB contributed to the development of GrittyWork, a digital workforce application and a registered trademark of the University of Colorado not related to this work. MB reports an honorarium from Med-IQ not related to this work. AP reports funding from a University of Colorado Innovations digiSPARK award not related to this work. AP contributed to the development of an AI-powered application not related to this work.

---

## References

1. Bagla P, Hanna J, Marthambadi B, Watkins S. Patterns of AI use in clinical work by hospitalists: survey study. *J Med Internet Res*. Mar 3, 2026;28:e85973. [doi: [10.2196/85973](https://doi.org/10.2196/85973)] [Medline: [41581898](https://pubmed.ncbi.nlm.nih.gov/41581898/)]
2. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. *JAMA*. Feb 8, 2022;327(6):521-522. [doi: [10.1001/jama.2021.25181](https://doi.org/10.1001/jama.2021.25181)] [Medline: [35061006](https://pubmed.ncbi.nlm.nih.gov/35061006/)]
3. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc*. Jul 2016;91(7):836-848. [doi: [10.1016/j.mayocp.2016.05.007](https://doi.org/10.1016/j.mayocp.2016.05.007)] [Medline: [27313121](https://pubmed.ncbi.nlm.nih.gov/27313121/)]
4. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open*. Oct 1, 2024;7(10):e2440969. [doi: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969)] [Medline: [39466245](https://pubmed.ncbi.nlm.nih.gov/39466245/)]
5. Everett SS, Bunning BJ, Jain P, et al. From tool to teammate in a randomized controlled trial of clinician-AI collaborative workflows for diagnosis. *NPJ Digit Med*. Mar 18, 2026. [doi: [10.1038/s41746-026-02545-1](https://doi.org/10.1038/s41746-026-02545-1)] [Medline: [41851268](https://pubmed.ncbi.nlm.nih.gov/41851268/)]
6. Rotenstein LS, Holmgren AJ, Thombley R, et al. Changes in clinician time expenditure and visit quantity with adoption of artificial intelligence-powered scribes: a multisite study. *JAMA*. Apr 28, 2026;335(16):1408-1417. [doi: [10.1001/jama.2026.2253](https://doi.org/10.1001/jama.2026.2253)] [Medline: [41920565](https://pubmed.ncbi.nlm.nih.gov/41920565/)]
7. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*. Sep 30, 2021;374:n2061. [doi: [10.1136/bmj.n2061](https://doi.org/10.1136/bmj.n2061)] [Medline: [34593508](https://pubmed.ncbi.nlm.nih.gov/34593508/)]
8. Glasgow RE, Trinkley KE, Ford B, Rabin BA. The application and evolution of the practical, robust implementation and sustainability model (PRISM): history and Innovations. *Glob Implement Res Appl*. 2024;4(4):404-420. [doi: [10.1007/s43477-024-00134-6](https://doi.org/10.1007/s43477-024-00134-6)] [Medline: [39568619](https://pubmed.ncbi.nlm.nih.gov/39568619/)]
9. Maw AM, Trinkley KE, Glasgow RE. The role of pragmatic implementation science methods in achieving equitable and effective use of artificial intelligence in healthcare. *J Gen Intern Med*. May 2024;39(7):1242-1244. [doi: [10.1007/s11606-023-08580-y](https://doi.org/10.1007/s11606-023-08580-y)] [Medline: [38172408](https://pubmed.ncbi.nlm.nih.gov/38172408/)]
10. Burden M, Dyrbye L. Evidence-based work design - bridging the divide. *N Engl J Med*. Mar 13, 2025;392(11):1044-1046. [doi: [10.1056/NEJMp2412389](https://doi.org/10.1056/NEJMp2412389)] [Medline: [40062670](https://pubmed.ncbi.nlm.nih.gov/40062670/)]

---

## Abbreviations

**AI:** artificial intelligence  
**EHR :** electronic health record  
**HIPAA:** Health Insurance Portability and Accountability Act  
**LLM:** large language model  
**PRISM:** pragmatic robust implementation and sustainability model

*Edited by Tiffany Leung; This is a non-peer-reviewed article; submitted 06.Apr.2026; final revised version received 24.Apr.2026; accepted 27.Apr.2026; published 01 Jun.2026*

*Please cite as:*

*Maw A, Pandita A, Burden M*

*Hospitalists Are Already Using AI—Why Implementation Will Determine Its Impact*

*J Med Internet Res 2026;28:e97419*

URL: <https://www.jmir.org/2026/1/e97419>

doi: [10.2196/97419](https://doi.org/10.2196/97419)

© Anna Maw, Aakriti Pandita, Marisha Burden. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 01.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.